5.

Building accessibility: platforms and methods for the development of digital editions and projects

Erica F. Cavanaugh and Jennifer E. Stertzer

Introduction

The twenty-first century has seen significant changes to the conceptualisation, creation and publication of digital editions. A long-standing practice, scholarly editing continues to thrive and evolve across many fields within the humanities. While print editions are still produced, the digital edition proves to be the most advantageous form to both editors and users. The combined presentation of digital editions created with structured data and project derivatives that repurpose source content provides a variety of mechanisms for which end users can discover, access and explore content. However, there are several challenges projects face when creating a digital edition, including the lack of comprehensive technical solutions, such as an approachable, flexible, powerful platform that supports all stages of digital editorial work and robust publication outputs.

This chapter will discuss the work of the University of Virginia Digital Publishing Cooperative (UVA-DPC), a grant-funded project with the goal of building the necessary infrastructure to facilitate and support the conceptualisation, development, publication, discovery, preservation and sustainability of digital editions and projects.[1] Major components

1 The UVA-DPC received generous funding from the Andrew W. Mellon Foundation.

of this work include addressing the critical issues, challenges, and opportunities that currently face the field of documentary editing: (1) the lack of accessible and robust digital editorial and publication platforms; (2) the issues of standardisation versus customisation, both within the editorial process and integrated into technical systems; (3) the need for diverse publication outputs; and (4) the immense potential for discoverability, increased accessibility and expansion of audience.

What is a digital scholarly edition?

The UVA-DPC broadly conceives 'publication' and developed the following definitions for publication outputs, including digital editions, digital derivatives and digital projects. The goal of the publication of a scholarly edition, whether print or digital, is to make historical documents accessible, both textually and intellectually. The first objective – accessible text – is accomplished by careful and professional transcription of the material. The second aim – intellectual accessibility – is achieved through scholarly apparatus such as footnotes, introductions, and essays, all written to decode the text (when was the manuscript written, when was it received, was any action taken, in what context was the letter written and identification of all the people, places, terms, quotations and so on). Most current projects, including those that have a print existence, are working towards a *digital edition*: a collection of historical documents that have been transcribed and edited following a consistent, transparent and well-informed editorial methodology and then published online. Because many of these editions were initially executed for a print publication, with digital versions being another instance of their print predecessors, it is important to incorporate traditional methods specific to the history of print documentary editions. These tools were devised to enable content discovery through scholarly methods such as indexes and document-specific annotation. It is important to note here that a 'digitised edition is not a digital edition'.[2] What distinguishes these

2 Patrick Sahle, What is a Scholarly Digital Edition? Pg. 27 https://www.jstor.org/stable/j.ctt1fzhh6v.6.

conversion projects – from print to electronic – lies in their use of the digital medium, thus bringing to bear the powerful organisation, search and display features of digital tools and platforms. These digital editions present the information contained in print editions as structured data, which enhances discoverability and makes re-usability possible. For example, a project may choose to create a cumulative index that integrates all legacy, volume-specific indexes into one resource allowing for easier access to the people, places, topics and events covered in an edition. Additionally, editors have the ability to present different 'views' of information, enabling users to browse documents by way of an edition's hierarchy (series, volume, page), chronologically or thematically.

It is important to add here that some of the more recent documentary editions are 'born-digital' editions for which there is no print version. Untethered from the protocols of paper, these works have begun to demonstrate how original digital publication facilitates creative processes for and presentations of the deep scholarly apparatus that has long distinguished the field of documentary editing. In this environment, projects have the ability to replace the traditional index structure with metadata and taxonomies. Annotations can also be re-imagined. In a born-digital edition, different types of annotation (document-specific, textual and general) can replace the footnote and endnote options of print.

Additionally, some editors present their findings in ways we call *digital derivatives*. Once online, projects may choose to make their document catalogues, their initial transcriptions, glossaries or their metadata (document, person, place) available online long before publishing a digital edition. These digital derivatives make available the outputs of editorial work during the process, thereby making historical and intellectual content accessible before an edition's scholarly editing and publication.

The term *digital projects* describes the web environment in which most digital editions and their derivatives exist. These ecosystems assemble the range of intellectual content created by an editor,

including blog posts, articles, data visualisations, timelines, presentations and so on that can be available alongside more traditional outputs. These opportunities provide editors with a variety of approaches to make content accessible and intelligible and appeal to large diverse audiences.

The three types of digital outputs – digital editions, digital derivatives and digital projects – allow for various modes of access and discoverability. While each output consists of intellectual content (transcriptions, annotations, indexes/taxonomies), metadata, interface and functionality, they offer different presentations of content (transcriptions, multiple annotation types, visualisations, data subsets) and paths of content discovery (searching, browsing, data exports). The rationale for this diversity is that different types of access appeal to different types of users. For example, many scholars are interested in transcriptions, annotations and indexes presented in ways that align with traditional print editions. On the other end of the spectrum, general audiences might be more comfortable exploring content through data visualisations and image-based icons that reveal specific information as well as present related content. At the core of both examples is well-structured content that enables this variety.

Why the UVA-DPC?

While tools and platforms developed by and for digital humanists have grown exponentially in the past few years, there is still no approachable, flexible, powerful platform that supports all stages of digital editorial work: (1) managing content (document catalogue, repository information and so on); (2) editing content (correspondence, diary entries, legal records and so on); (3) capturing information about that content (metadata); (4) providing context (annotation, indexes, data visualisations, metadata and so on); (5) tracking workflow steps and versions of content; (6) and, making all content accessible (by way of various forms of digital publication). Furthermore, finite resources, including the availability and afforda-

bility of technical expertise, and the lack of stand-alone platform solutions and digital publication options, limit what editors can produce.[3] Unlike other workflows that depend on multiple digital tools and platforms to produce digital publication outputs,[4] our goal at the Center for Digital Editing has been to help solve these challenges by creating a single, comprehensive, flexible system. Over the years, we have created several all-inclusive editorial systems for our partner projects that support all stages of editorial and publication work, allowing the editor to work within one environment and produce one or more digital publication output, all using Drupal.

An open-source content management system, Drupal allows users to build highly customised sites to capture complex data that is often seen in digital humanities projects. It is flexible, has a large user community and 'allows scholars with a much lower level of comfort with technology to build much more complex projects'.[5] However, while it is a powerful tool, creating a Drupal-based site on your own can be daunting. Its customisability is both its strength and weakness, forcing users to traverse a notoriously steep and rocky learning curve to develop their site while also taking on the task of data modelling, one of the most important aspects of developing any site using

3 During the planning year, the UVA-DPC performed an environmental scan of technologies in use in the field of documentary editing. We also evaluated the different technologies and workflows in use at the participating projects.

4 One example of a project that utilises this approach is the Papers of George Washington: editorial work takes place in MS Word; the Word file is styled (for typesetting purposes) before being sent to the press; once at the press, the Word file is converted to .pdf for print publication; for digital publication, the .pdf file is converted to XML and uploaded to a version-control repository for additional encoding and the integration of single-volume index file into the cumulative index files; source XML files are transformed for publication on Rotunda and Founders Online platforms. Additionally, the project maintains its document catalogue in Drupal, digital resources in Google sites, and its project website in WordPress..

5 Quinn Dombrowski, https://quinndombrowski.com/blog/2019/11/08/sorry-all-drupal-reflections-3rd-anniversary-drupal-humanists/.

Drupal.[6] But what if users were able to bypass the rocky slope and the majority of data modelling altogether?

The UVA-DPC has been working to develop a Drupal-based module that would remove these hurdles for future editorial projects. While heavily influenced by the work at the Center for Digital Editing with projects like the George Washington Financial Papers Project and the Papers of Martin Van Buren, the UVA-DPC technical team collaborated with editors and editorial projects that are thematically, chronologically and methodologically both diverse and similar. These collaborations provided a deeper understanding of how documents are structured and formated, how they relate to one another and how best to craft annotations and indexes that allow for intellectual accessibility. As a result, this platform supports the major workflow components of creating digital publication outputs – content management, editorial process, content markup and visualisation, and user interface development – while adding a layer of standardisation including name authority files, metadata standards, shared vocabularies, and data models. Drupal-based digital editions which previously took months to create the underlying infrastructure can be created in just a few minutes using this module's prebuilt templates to capture various document elements. This isn't to say data modelling should be ignored, but instead that editors won't be starting with a blank slate, scratching their heads and wondering where to start. Projects will be able to make use of a series of standardised fields and layouts to capture information. If they then determine that additional fields are necessary, projects will be able to add them, which would require both some data modelling and tackling the notorious learning curve. But instead of a treacherous mountain, the module that we are creating will help to make it more like a large hill.

6 Quinn Dombrowski, https://drupal.forhumanists.org/drupal-humanists-chapter-1-first-things/when-use-drupal.

Module components and features

The UVA-DPC module is built on top of a collection of Drupal modules and the configuration settings that we have designated for them. During development, the UVA-DPC tech team placed a strong emphasis on using well-supported and established modules to create the desired functionalities needed for the primary goals of digital editions. There has been a strong focus on limiting the amount of custom coding and steering clear of features that had little support within the Drupal community. By following a 'less is more' philosophy, the sustainability of the module and those sites using it markedly increases.

Upon installation, the module will add a series of content types (templates for collecting data) and all associated fields. The fields related to the document/object content type are extensive, covering the various phases of workflow: cataloguing, organisation, selection, transcription, annotation, publication and any steps in between. These fields can put documents in conversation with each other directly via relational fields, such as responses or enclosures, as well as indirectly using keywords and descriptive metadata. To encourage standards used within the scholarly editing community, the fields created can be mapped directly to specific elements in TEI XML. Another built-in content type captures information about people mentioned within an edition. The fields for the person content type correlate to several elements from the Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF).[7] Furthermore, fields have been included to encourage the use of authority records when possible. The inclusion of individual fields for the Virtual International Authority File (VIAF), Wikidata and Enslaved: Peoples of the Historical Slave Trade[8] within the platform familiarises the

7 See https://eac.staatsbibliothek-berlin.de/.

8 The Virtual International Authority File (https://viaf.org) brings together multiple name authority files into one. Enslaved: Peoples of the Historical Slave Trade (https://enslaved.org/) is a discovery hub documenting people of the Atlantic slave trade, both enslaved and freed. Many of these individuals would not have identification in data sets like VIAF.

editors with these options and encourages the use of authority records early on in the editorial process. These options serve as possible starting points; knowing that there are a number of authority records available for use, the platform also allows editors to customise by adding additional authority records as necessary.

Document transcription and the representation of the text as it appears in the manuscript are important features for many editions. The editorial interface allows editors to transcribe each document directly into the system. The transcription fields use a standard WYSIWYG editor with markup buttons that most editors are accustomed to seeing – bold, italics, strikethrough and so on. The options have been expanded to allow for a series of custom styles, including marking paragraph indents, small caps, hanging indents and so on. Additionally, the insertion of various forms of media into these fields is also possible, along with a robust symbols list that has been adapted to include Unicode-based symbols that are often seen in historical manuscripts.

Creating layouts and page displays that directly call on the data and transcriptions entered into the system can be a difficult task for editors to first get the hang of. To facilitate this work, the module includes a series of preconfigured views and blocks[9] that can be cloned and altered by individual projects. These layouts include some of the digital derivatives mentioned previously; for example, a searchable document catalogue and glossaries for all people and organisational records added to the digital edition. Other preconfigured content includes a .pdf download link of a published record, social sharing links to aid in content sharing and a standard search interface. While other views and blocks are available, the UVA-DPC technical team was cognisant of the fact that it is impossible to anticipate the way in which all projects using the module would want to call on their content. However, through the options already avail-

9 Views can be understood as SQL-queries done through a user interface to display site content in various formats. Blocks are pieces of content that can be rendered onto various regions of a page.

able and comprehensive documentation, editors will be able to create additional layouts as needed.

As this module is focused on allowing editors to quickly create digital editions with best practices in mind, there will be a number of added features and functionalities. Customised user roles and project workflows to help track productivity are current features of this module. More importantly, however, is the potential to preconfigure taxonomy lists so that all projects can use consistent terminology that has been mapped to existing controlled vocabularies. The benefits of using shared ontologies include the increased potential for interoperability and cross-site searching, resulting in increased discoverability. Some taxonomy lists can be easily shared across a large and diverse corpus of projects, such as language and document/object type. Whereas others, such as keywords, are significantly more challenging. Numerous conversations between members of the UVA-DPC and editors working on digital editions have taken place to gather feedback. As a solution, two fields have been added to capture keyword information: a UVA-DPC-controlled 'theme' taxonomy list and a project-specific 'keyword' taxonomy list. This allows for a standard set of general terms to be used for potential cross-site searching while allowing projects to have the flexibility on more specific terminology used within their edition. The cooperative will provide guidance and resources for controlled vocabularies projects might consider, but we are also aware that these vocabularies can often fall short when it comes to capturing the history, culture and experiences of diverse communities.

Projects at the beginning phases will be able to use the platform by simply installing the module on a Drupal site and begin adding content. As a web-based application, multiple project staff members can take laptops into an archive and simultaneously catalogue records. However, some projects will enter the process with inprogress or completed editorial work. In these instances, examples are included on how to import content from CSV or XML file formats. Realising that it is impossible to anticipate the exact state content will be in when projects decide to import content into the platform,

the built-in importers also serve as examples that, used in conjunction with documentation, can be duplicated and adjusted as needed. Alongside the ability to import content is the ability to export content into stable, structured data formats (CSV, XML and JSON). This feature not only allows the users and the editors to download content (if enabled) but also makes it possible for hybrid publishing options via both Rotunda[10] and Drupal, as well as the ability to send content from the platform to GAMS.[11]

Since the platform is able to generate export files in CSV formats and supports the creation of REST APIs, the UVA-DPC technical team has explored the option of treating collections as data instead of simply as a static representation of materials to be found on a website.[12] Configurations are included upon module installation for file paths to export published content to be used for archival storage. In collaboration with the University of Virginia Libraries, projects within the cooperative will be able to add their data sets to the institution-based archival digital storage solution, LibraData.[13] All projects adding materials can record information on collaborators and contributors, funders, keywords, project history, editorial methodology and so on, and receive a DOI upon initial creation. Moreover, upon publication, these data sets become discoverable through UVA library's main catalogue, Virgo. Taking advantage of this archiving resource allows projects, especially new projects, to have a clear

10 The digital imprint of the University of Virginia Press. The UVA-DPC includes an option to publish content in Rotunda, following the Press's submission process.

11 An Open Archival Information compliant asset management system for the management, publication and long-term archiving of digital resources. Tabulare financial records are the primary forms of content being sent to the GAMS repository.

12 Thomas G. Picadilla, Collections as data: Implications for enclosure. Association of College & Research Libraries, vol 79, no 6, 2018.

13 Libra makes publications and data sets freely available and provides storage through the Academic Preservation Trust. Libra is part of UVA Libraries and is available for use by UVA-affiliated individuals. For more information, see https://www.library.virginia.edu/libra/.

long-term preservation and sustainability plan for the content with which they are working, even when the technology used for the digital edition may no longer be available or functioning.

Who is the module for?

Two user groups stand to benefit from the aforementioned work: editors/editorial projects and end users/project audience. First, simply put, the UVA-DPC will offer editors optimal paths to digital dissemination at various stages of their workflow, from digitisation to publication. These editors include not only scholars/academics, but those some would refer to as 'accidental editors'[14] who do not readily identify themselves as editors. The UVA-DPC will have the capacity to work with projects from conceptualisation through publication, a process that will inevitably vary depending on the project's needs.[15] The work of developing solutions (and discovering pain points along the way) for the projects currently part of the cooperative – and the diversity they represent – will ensure this process is comprehensive and will help the UVA-DPC prepare to accept new projects.

The digital outputs from the UVA-DPC module will also benefit the end users of digital publications. Audiences for documentary editions have expanded beyond scholars, as more editions are available online. Students, teachers, general readers and genealogists, as well as organisations and individuals involved in the repurposing of content (including lesson plans, social media content and primary source-based web content) seek access to primary source materials. This broader audience inspires and requires editors to employ strategies and develop digital outputs to increase that accessibility even more.

14 Accidental Editors, Ben Brumfield, https://www.sidestone.com/openaccess/9789088904837.pdf.

15 The CDE developed a model for this type of partnership and has successfully worked with projects over the past three years in this capacity. Additionally, staff from both the CDE and the University of Virginia Press have a long history of working together to publish digital editions.

The diversity of digital outputs is essential if we are to engage with and expand our audiences. But what does this mean for both the current and future projects that publish via the UVA-DPC? How will this work and collaboration benefit end users and advance research and understanding? First and foremost, by working with the UVA-DPC, editors will be able to conceptualise their projects and publications in ways that align with their accessibility goals. For instance, an editor might want to make their editorial process transparent and make their document catalogue, first-pass transcriptions and metadata available long before they publish their digital edition.[16] Another example involves several projects within the UVA-DPC that share thematic connections. A federated platform is being created to merge content from the Drupal-based editions so that users may search multiple projects at once. Editors could also choose to develop a digital project, such as the George Washington Financial Papers Project, that incorporates an interactive e-book to help users understand the site, articles from other historians to help contextualise the content, data visualisations, project timelines, and conference poster/paper presentations, alongside the project's digital edition of complex financial records.[17] In short, providing multiple paths of discovery (by way of different digital outputs), federating thematically and chronologically similar content, and developing numerous outputs will increase accessibility and advance understanding.

16 For examples, see The Papers of Martin Van Buren (http://vanburenpapers.org) and The Papers of Julian Bond (https://bondpapersproject.org). Documents from both of these sites will eventually be fully edited and made available in digital editions.

17 See the George Washington Financial Papers Project's site (http://financial.gwpapers.org).